ARMY RESEARCH LABORATORY

Progress Toward Adaptive Integration and Optimization of Automated and Neural Processing Systems: Establishing Neural and Behavioral Benchmarks of Optimized Performance

by Anthony J Ries, Laurie Gibson, Jon Touryan, Kaleb McDowell, Hubert Cecotii, and Barry Giesbrecht

ARL-TR-7147 **November 2014**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5425

ARL-TR-7147 **November 2014**

Progress Toward Adaptive Integration and Optimization of Automated and Neural Processing Systems: Establishing Neural and Behavioral Benchmarks of Optimized Performance

Anthony J Ries, Jon Touryan, and Kaleb McDowell
Human Research and Engineering Directorate, ARL

Laurie Gibson
Science Applications International Corporation
Louisville, CO

Hubert Cecotii, and Barry Giesbrecht
University of California at Santa Barbara
Santa Barbara, CA

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
November 2014	Final	August 2012–August 2013

4. TITLE AND SUBTITLE
Progress Toward Adaptive Integration and Optimization of Automated and Neural Processing Systems: Establishing Neural and Behavioral Benchmarks of Optimized Performance

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Anthony J Ries, Laurie Gibson, Jon Touryan, Kaleb McDowell, Hubert Cecotii, and Barry Giesbrecht

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Research Laboratory
ATTN: RDRL-HRS C
Aberdeen Proving Ground, MD 21005-5425

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-7147

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Technical advances intended to improve situational awareness by providing more information about the tactical environment place high demands on the Soldier's limited capacity cognitive and neural systems. Information display technologies have been developed that filter information to prevent performance failures due to information overload. However, these technologies are typically rigid with respect to changes in the operator's physical and cognitive state. The objective of the project described in this report is to develop an adaptive framework that adjusts filtering algorithms to optimize human performance in a variety of operational contexts. The work adopts a unique approach that integrates measures of behavior and brain activity with automated information processing and display algorithms. It leverages basic science research conducted at the Institute for Collaborative Biotechnologies that uses machine learning algorithms to detect performance failures during difficult attentional tasks based on brain activity, work done at Science Applications International Corporation using pattern classification algorithms to detect threats based on brain activity, and work done at the US Army Research Laboratory aimed at understanding the cognitive constraints on performance in crew stations.

15. SUBJECT TERMS
rapid serial visual presentation, RSVP, EEG, neural classification, P300, brain-computer interface

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Anthony J Ries
Unclassified	Unclassified	Unclassified	UU	42	19b. TELEPHONE NUMBER (*Include area code*) 410-278-0915

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contacts

List of Figures

List of Tables

INTENTIONALLY LEFT BLANK.

1. Introduction

With advances in display and sensor technologies and with increased emphasis on a smaller, more mobile fighting force, today's Soldier must deal with a density and complexity of information that was unknown in the past. Although the intent of providing Soldiers with more information is to improve their situational awareness and operational performance in tactical situations, the increased informational content places high demands on limited-capacity cognitive and neural systems. Various automated filtering algorithms and adaptive displays have been developed to help reduce the amount of information presented to the Soldier, but these algorithms are rigid and do not adjust based on the user's cognitive capacity, strategies, and level of stress. As a result of this rigidity, their inefficacy can result in suboptimal use of information. Ultimately, even with high-performance automated filtering systems, the burden is on the Soldier to act on the information in dynamic, complex environments. Therefore, it is critical to develop technologies that will allow the integrated human-machine system to be highly adaptive to any context.

This report documents the results of the second year of a 3-year project to develop an approach for integrating measures of neural activity into complex multiplatform human-machine systems that will provide real-time classification of cognitive and perceptual states and will provide dynamic, adaptive adjustment of information displays to accommodate fluctuations in these states. The project builds upon key basic research conducted at the Institute for Collaborative Biotechnology, applying measures of brain activity to classify performance failures during difficult attentional tasks. The overall goal of the project is to establish fundamental parameters for optimizing attentional state classification in dynamic tasks from measures of brain activity. These measures will be integrated with other measures of behavioral performance and physiology and instantiated in hardware and software to monitor and optimize Soldier performance.

In year 1, the team developed benchmarks and studied key display parameters for operator performance within demanding tasks. The year 2 work was focused on translating what was learned during year 1 (see Gibson et al. 2012) into more-realistic multitasking environments. Behind the work are basic questions about the utility of rapid serial visual presentation (RSVP) and neural signal processing compared with more conventional interface paradigms. What are the types of systems and tasks for which brain-computer interfaces (BCIs) provide the biggest improvement in performance? Can we demonstrate clear advantages with BCIs?

2. Simulator Development and Experiments

In addition to communications tasks and to tracking the vehicle position on a mission map, the manned ground vehicle (MGV) commander maintains situational awareness of the environment outside the vehicle. This is done by viewing a 180° field of view (FOV) and by scanning the vehicle surroundings with a 3-axis pan-tilt-zoom camera (PTZ). We will refer to this scanning task as "portal search". The commander may also be cued to critical events through an auditory processing system that can detect the approximate location of the gunshots and explosions. In this case, the viewing portal will be automatically moved to the approximate location of the sound source ("slewed to cue") to augment target visual search. Figure 1 shows a display from the US Army Tank Automotive Research, Development and Engineering Center (TARDEC) simulator with the controllable portal in the lower left quadrant. We identified the portal search task as one that could potentially be replaced by intelligent RSVP where a computer algorithm searches imagery from the immediate vehicle surroundings for salient objects or regions of interest (ROI) and presents only those images to the human operator. The neural response of the commander elicited by these stimuli is processed by machine learning algorithms to identify likely targets among the many distracters. In other words, instead of the commander manually searching with a joystick, he simply views images of the surroundings that have been identified as possibly containing targets or threats.

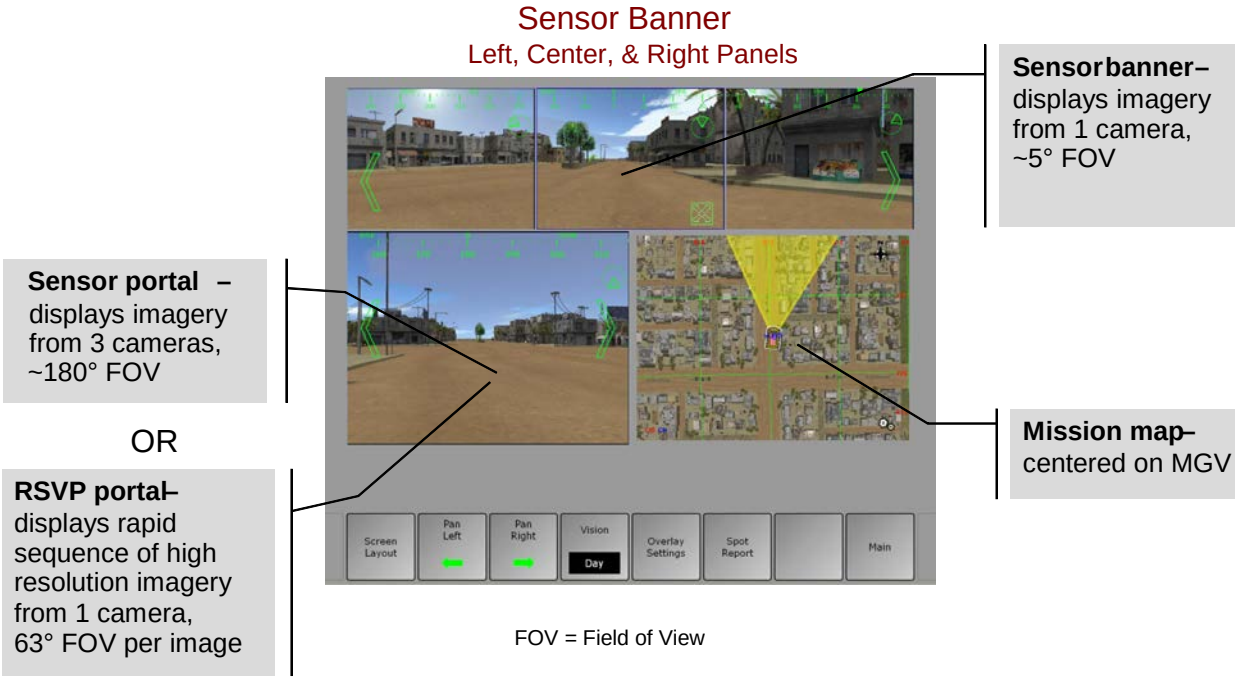

Fig. 1 Display from the TARDEC simulator of a crew station commander's view. The bottom left quadrant is currently a controllable portal that could be replaced by intelligent RSVP.

A key question about this use of neural processing is whether or not it would improve overall performance. Clearly, the answer to the question depends on how efficient and accurate the manual search is, how well the image filtering algorithm works, and how quick and accurate the RSVP process is. Specifically we identified the parameters in Tables 1 and 2 to quantify portal search and RSVP performance for the purpose of making comparisons.

Table 1 Manipulated parameters

Variable	Conditions	Purpose
Portal speed	1. Slow 2. Medium 3. Fast	Portal speed should directly affect search time. However, this parameter may be fixed or constrained by the hardware (e.g., PTZ sensor).
Portal accuracy	1. Near target 2. Far from target	The slew-to-cue accuracy affects the initial placement of the camera/portal relative to the target and should have a large impact on search speed.
Portal context	1. No 2. Yes	Portal or scene context will directly influence the search strategy and thus search time.
RSVP false alarm rate	1. Low (10:1) 2. High (100:1)	The false alarm rate will increase total RSVP processing time and may impact behavioral accuracy.
Target salience	1. Low[a] 2. High[a]	Salience will determine visibility constraints on observer and classifier performance.

[a]Quantified by established model of visual saliency (Itti and Koch 2000)

Table 2 Fixed parameters

Variable	Value	Purpose
RSVP presentation rate	2 Hz	RSVP presentation rate will be fixed to the standard rate (2 Hz) employed by the neural processing system.
RSVP and portal size	300 × 300 pixels	Since the purpose of these experiments is to compare RSVP and portal search, the absolute size is irrelevant.

With these parameters we defined a series of experiments to examine the tasks and conditions under which RSVP is beneficial. The goal was to determine the parameter space for which RSVP improves detection performance (speed and accuracy) relative to the portal search (i.e., direct control of the PTZ). The initial experiments collected only behavioral data since the performance of the classifiers used to discriminate neural signals, under single-task RSVP conditions, is well known (Touryan et al. 2010). Later, using a more complex multitasking simulator with integrated real-time electroencephalogram (EEG) processing, RSVP performance was measured. Figure 2a shows the display for the portal search component of the first experiments. Here, stimuli were generated from the video game “Call of Duty: Black Ops” (Activision, Santa Monica, CA, 2011). Targets are dismounts with guns and are either present or not in the images.

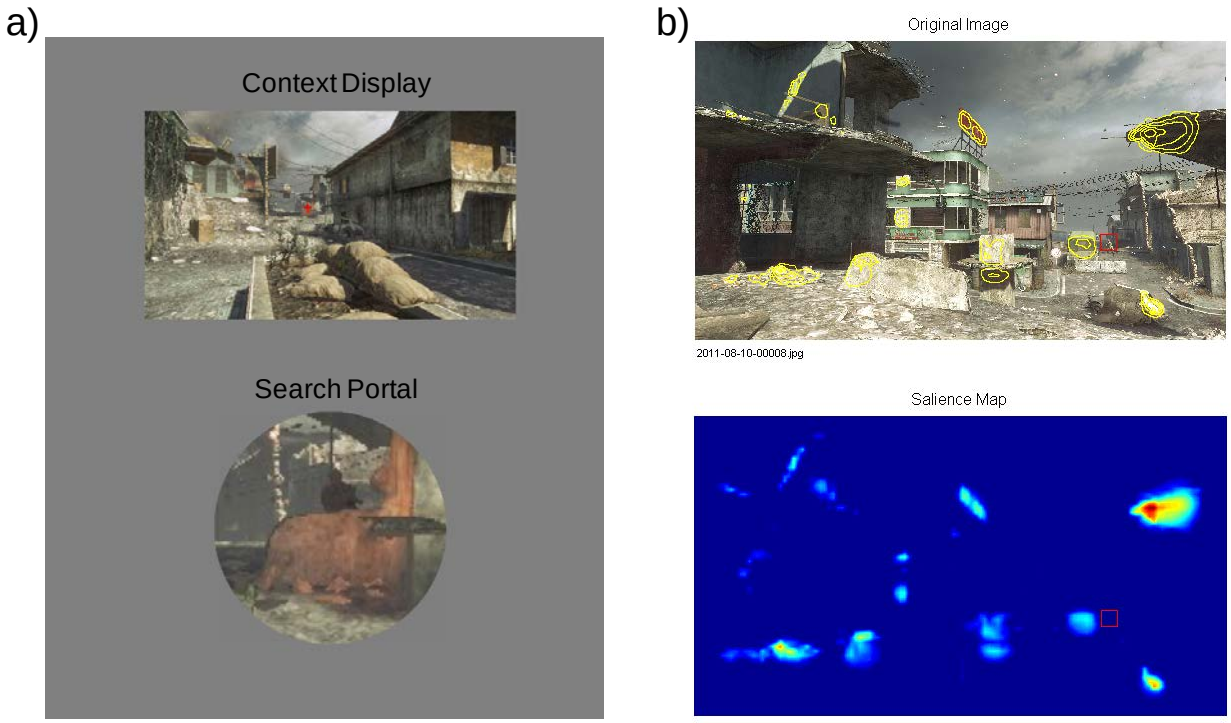

Fig. 2 Portal search and salience: a) portal search task display; top window is the context display and the bottom window is the controllable portal (centered on target) and b) example stimuli with salient features outlined (top image) and colorized (bottom image). The target (contained within the red box) does not register as a salient feature.

To compare manual versus RSVP performance, it is important to understand how an automated filtering algorithm would integrate with and affect RSVP. For example, the number of false alarms generated by the filtering algorithm directly impacts the length of the RSVP sequence, i.e., the number of images that must be presented to the operator. For the purpose of this evaluation, we considered an established technique for identifying salient features within a natural scene (Itti and Koch 2000) to use for automated image filtering. Figure 2b illustrates salient features and objects identified with this algorithm in a sample scene. While many of the portions of this image are salient in terms of contrast, orientation, and color features, they are not, unfortunately, the portions that include the target object. This is not unusual. In this context targets are typically occluded or camouflaged as would be expected in an operational military environment. Thus the saliency approach based on bottom-up features to automated filtering is not particularly useful here. Rather, the optimal algorithm would have to incorporate contextual information and be general enough to detect many types of targets or objects of interest (e.g., people, vehicles, guns, windows) without a large number of false alarms.

For our comparison studies we decided to emulate an automated filtering algorithm and directly control the RSVP false alarm rate by manually identifying objects of interest (including the target) in each of the images. We set the ratio of images presented during RSVP that contained a target to the images that contained no target and address the question of the efficacy of RSVP

under conditions that range from target-sparse to target-rich environments. By emulating the algorithm, we could meet the goal of understanding the conditions under which replacing the manual search with prefiltered RSVP leads to improved performance.

Three different types of studies were conducted in support of the development of the simulator. The first study looked at search times and accuracies for a manual portal search compared with automated filtering and RSVP. The second study focused on using RSVP for threat detection and building neural response models that could be used to automatically (versus manually) indicate when the operator detects a threat based on his EEG signals. In the third study, participants carried out both portal search and the RSVP search with classification of their neural signals, all within the multitasking simulation environment. These studies and the results are described in the following subsections.

2.1 Study 1: Portal Search Experiment

In testing with outside subjects, initial results indicated that the portal task was a good paradigm for quantifying the effects of the relevant parameters on performance. In this task subjects alternate between portal search and RSVP. In the portal search blocks they must move the portal (PTZ) until they find the target or decide that there is no target present. The initial placement of the portal is randomly distributed within a given window around the target. In most cases the target cannot be seen in the context display and must be identified within the search portal. The context display serves primarily to influence the subject's search path and provided information on likely target locations (doors, windows, cars, etc). Figures 3 and 4 illustrate search paths from 2 subjects. While the portal was initially placed near the target, the subject chose to move the portal along a search path away from the true target, resulting in a relatively long search time.

Fig. 3 Example of a search path. The yellow "+" indicates the initial placement of the portal while the red "x" indicates the final placement and target detection. Inset shows the final portal image containing the target.

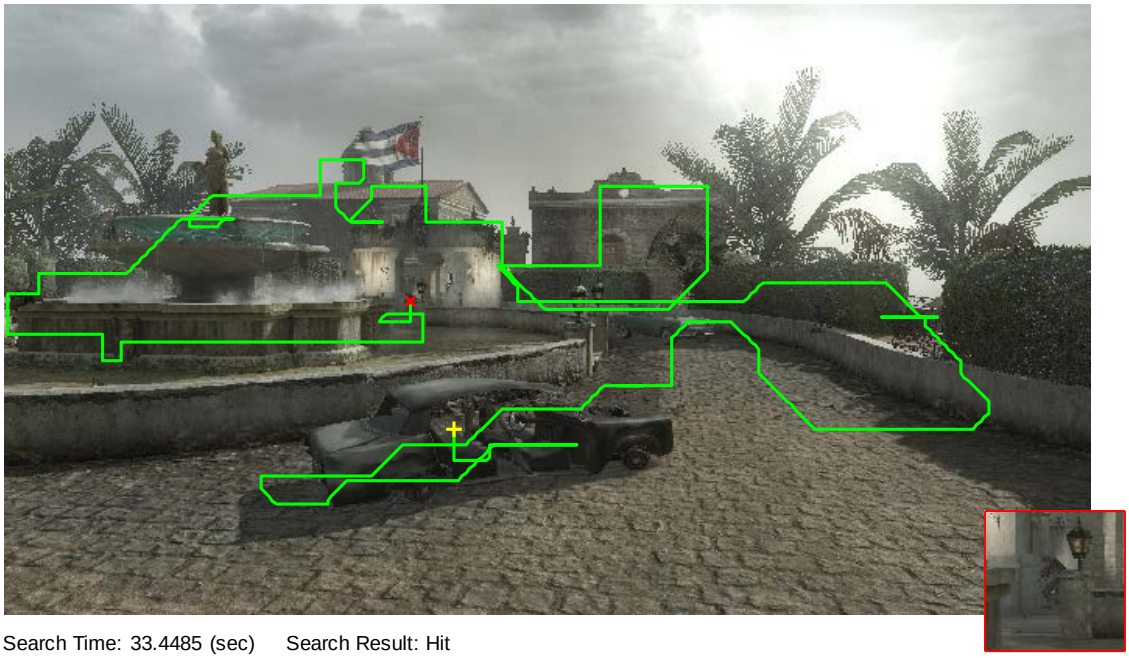

Fig. 4 Example of a search path. The yellow "+" indicates the initial placement of the portal while the red "x" indicates the final placement and target detection. Inset shows the final portal image containing the target.

Data from 16 subjects in the portal search experiment are shown in Fig. 5. The top panels show the search time distributions for target-present (red) and target-absent (blue). Vertical lines indicate distribution mean. The lower left quadrant shows the relationship between initial portal placement and search time. The lower right quadrant shows the time to target in the RSVP condition with a 10-ROI sequence (9 false alarms, 1 target). As expected, the search time for target-present images was significantly shorter that for target-absent ($p < 0.001$, Wilcoxon rank sum test). The average search time for target-present images was 18 s while the average search time for target-absent images was 60 s. While the accuracy for the portal search component was high (mean total accuracy = 0.85), it was substantially lower than for the RSVP component (mean total accuracy = 0.99). There was a significant correlation in the accuracy between subjects' manual search and RSVP performance ($r = 0.62$, $p = 0.01$). Because of the high accuracy in the RSVP component (identified in the preliminary studies), we decided to keep the RSVP false alarm rate fixed at 0.1, i.e., 10 nontarget ROIs (image clips) for each target ROI in all subsequent experiments. Under these conditions the false alarm rate could be tripled (to 0.3) and still outperform the manual search.

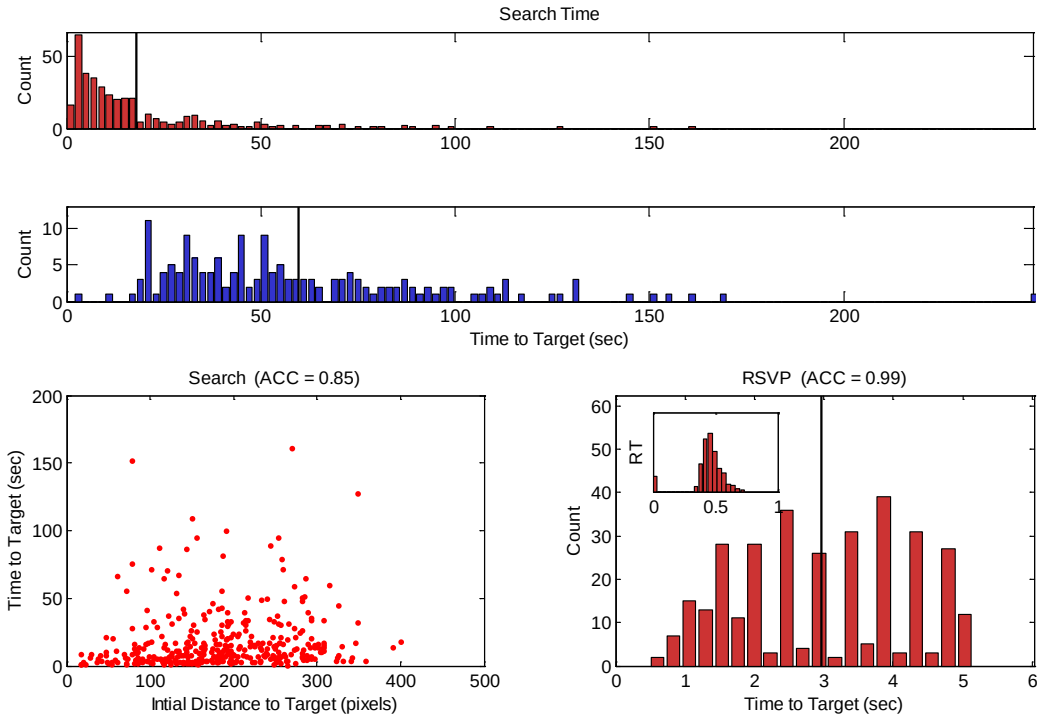

Fig. 5 Portal search and RSVP summary statistics (16 subjects)

One of the most interesting observations from this experiment is that the search time does not strongly correlate with the initial portal accuracy (i.e., distance from the portal center to the target). The correlation coefficient between search time and accuracy is 0.09 ($p = 0.07$). Unless the portal is placed within 100 pixels of the target, the initial placement of the portal does not influence the search time. This observation speaks directly to the importance of the slew-to-cue

accuracy and the training of the operator. If the slew-to-cue accuracy can be well quantified, it will be imperative to instruct operators to stay within the area of initial placement and suppress their instinct to follow a contextual search path. In a similar fashion, the intelligent RSVP should be programmed to give priority to ROIs that fall within the cued area.

One of the other key parameters in Table 2 is portal speed. To test this directly we manipulated portal speed for a subset of subjects (N = 10). These subjects performed the search experiment in 2 sessions. In each session their portal speed (in PTZ) was set to a value of either baseline (1× condition) or twice baseline (2× condition). Half of the subjects had the 1× condition first and half had the 2× condition first. Over the population we found that there was no significant difference in search time for the 2 conditions ($p > 0.05$, Wilcoxon rank sum test). However, we did find a significant reduction in search time between session 1 and 2, indicating a practice effect ($p < 0.05$, Wilcoxon rank sum test). These results suggest that training rather than gimbal speed is more important for system performance.

Finally, we quantified the relationship between target salience (Itti and Koch 2000) and search time. As expected, there was no significant correlation between target salience and search time ($r = 0.03$, $p = 0.51$). This is primarily due to the fact that the majority of targets were low salience. When the targets do not "pop out" of the background, subjects follow a search path guided by the large-scale contextual cues (e.g., buildings, cars, doors, windows). As such, the automated filtering algorithm should incorporate contextual cues and not just low level feature salience (Torralba et al. 2003; Torralba et al. 2006).

2.2 Study 2: RSVP Experiment 1

In addition to the portal search study, a subset of subjects (N = 12) also performed a longer RSVP experiment. Here the stimulus set was from the same ensemble as the portal search experiment and the RSVP presentation rate was fixed at 2 Hz. However, in this experiment subjects viewed 10 blocks of RSVP, each 2 min long. EEG recordings were digitally sampled at 256 Hz from 20 scalp electrodes, located on the standard 10–20 coordinate grid, using an Advanced Brain Monitoring (ABM) ×24 system configured with the single-trial event-related potential (ERP) sensor strip and operating in wired mode (Advanced Brain Monitoring, Carlsbad, CA). While the headset operates in both wireless and wired modes, the wired mode provided the best event timing, which is critical in RSVP experiments.

The 12 subjects indicated when they saw a target during the RSVP sequences by pressing a key. Their mean accuracy for detecting targets was 0.86 (minimum = 0.73, maximum = 0.94, $\sigma = 0.06$). This data is shown in Fig. 6. The EEG data for each subject was used to create a model of the subject's response to a target for the ABM headset. The models were constructed using a machine learning algorithm described elsewhere (Touryan et al. 2010). The mean area under the receiver operating characteristic (ROC) curve for the 12 models was 0.93 (minimum = 0.81, maximum = 0.97, $\sigma = 0.04$). Figure 7 presents the data for each subject. Blue bars represent the area under the ROC curve for the classifier using models built from the RSVP

session and applied for each of the 12 subjects. The green bars represent the area under the ROC curve for 2 subjects when these models are applied to the data from a second session, at least one week later. As one might expect, the accuracy of the classifier models is highly correlated with target detection accuracy ($r = 0.71$, $p < 0.01$). Figure 8 shows the classifier model that was built for subject 104.

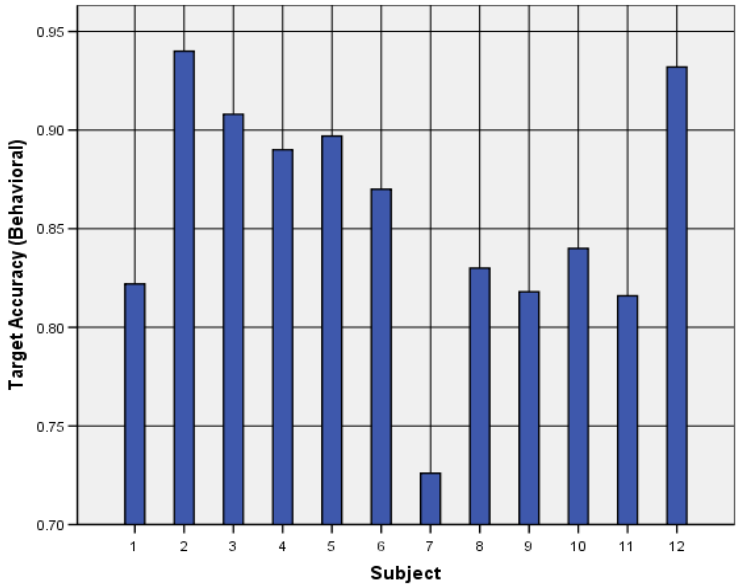

Fig. 6 Accuracy in detecting targets as recorded with a key press for the 12 subjects in the RSVP study

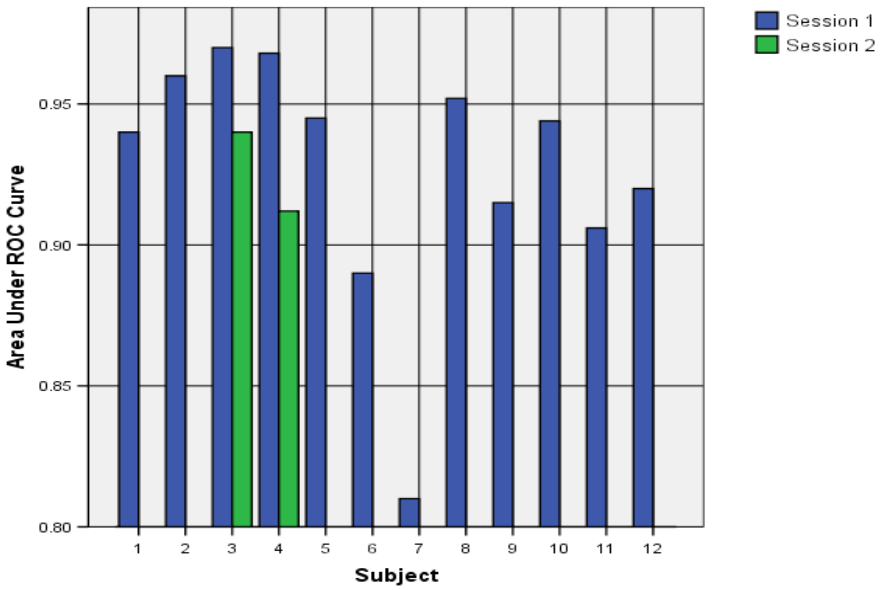

Fig. 7 Subject-by-subject data

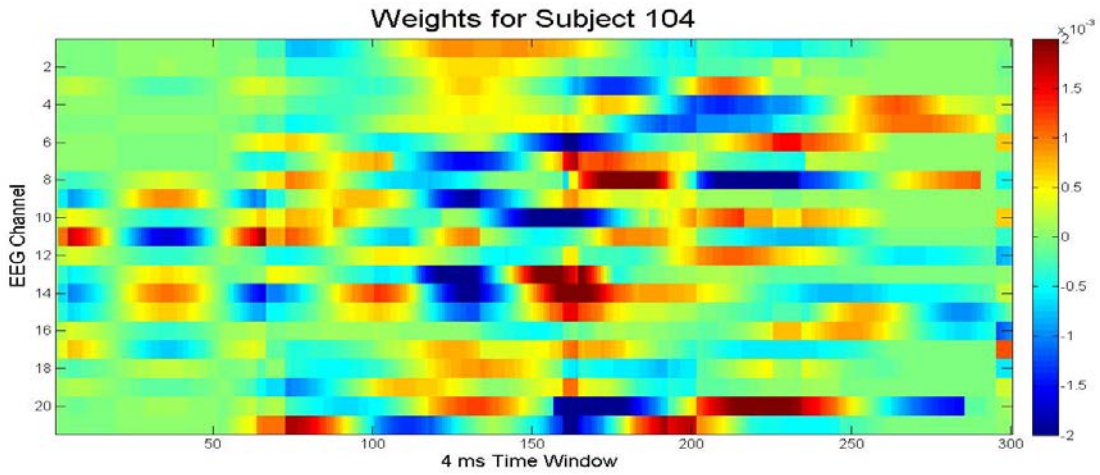

Fig. 8 Graphical representation of the classifier weights for a single subject

Previous studies have shown that accurate target detection in RSVP can be maintained at significantly higher presentation rates (Sajda et al. 2003, Luo and Sajda 2009, Sajda et al. 2010). Since an increase in the presentation rate would substantially reduce the search time, we modified the RSVP rate to 5 Hz. To test the performance of the simulator, we again needed to build models for each subject using the new 5-Hz presentation rate.

2.3 Study 3: RSVP Experiment 2

In this experiment we used the same RSVP paradigm and EEG acquisition system as the previous experiment. Subjects viewed a stream of images (this time presented at 5 Hz) that were cropped screenshots from the video game "Call of Duty". A small percentage of these images (~5%) contained the target: a Soldier carrying a gun. Subjects were instructed to respond with a button press when they saw the target. Since each frame was only displayed for 200 ms, the button response typically occurred well after the target image was presented. To compensate for this response lag, we used a heuristic method to assign each button press to a particular image. Specifically, for each button press we identified the preceding 3 images. If one of those images was a target, the response was assigned to that image. If none of those images was a target, the incorrect response (false alarm) was assigned to the image corresponding to the average reaction time (~500 ms). Using this method we were able to get reasonable estimates of reaction time and accuracy.

To quantify the behavioral performance we used the F-measure, which incorporates both detection and false alarm rates (Fawcett 2006). Specifically, the F-measure combines the precision (positive predictive palue, *PPV*) and hit rate (true positive rate, *TPR*):

$$PPV = \frac{TP}{TP + FP}, \tag{1}$$

$$TPR = \frac{TP}{TP + FN}, \tag{2}$$

and

$$F\ measure = 2\frac{PPV * TPR}{PPV + TPR}, \tag{3}$$

where *TP*, *FP*, and *FN* are the number of true positives, false positives, and false negatives, respectively. Over the population of 13 subjects, the average *F-measure* was 0.81 (minimum = 0.692, maximum = 0.909, σ = 0.069). Likewise, to quantify the classifier performance, we again used the area under the ROC curve (AUC). For the 13 subjects, the average AUC was 0.875 (minimum = 0.701, maximum = 0.975, σ = 0.069). Figure 9 shows the relationship between the *F-measure* and the AUC. As in the 2-Hz condition, there was again a significant correlation between the behavioral and classifier performance ($r = 0.618$, $p < 0.05$).

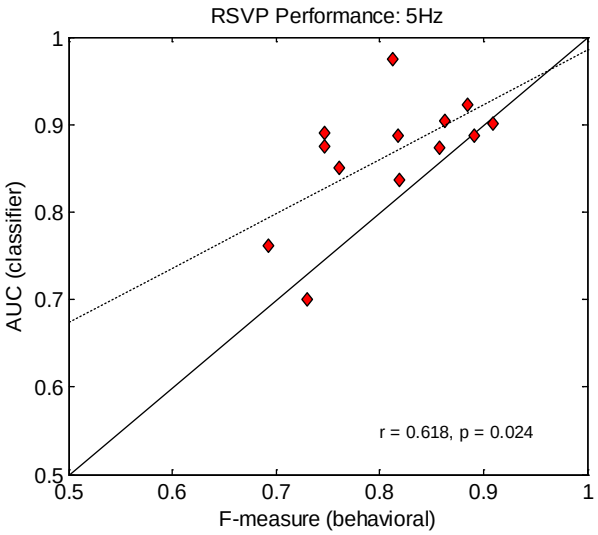

Fig. 9 Behavioral and classifier performance in the 5-Hz RSVP condition

In addition to analyzing the performance of the classifier, we examined the relationship between the classifier score and stimulus properties. Specifically, does the score indicate the perceptual difficulty of the target detection? In the RSVP stimulus ensemble there is a range in target visibility. In some target images the Soldier is large (occupying up to one-third of the image) and salient. In other target images the Soldier is small and distant or partially occluded (or both). Likewise, strong shadows and camouflage prevent some targets from being easily distinguished. One way to quantify the visibility would be through a parameterization of the target and background pixels (e.g., average salience, faction of image, luminance, and contrast). However, an aggregate behavioral response provides a quick proxy for target visibility. Here we calculated the average hit rate for each target across all subjects (N = 13) in the 5-Hz RSVP experiment. The increased RSVP speed resulted in a more-sensitive measure of target visibility with average hit rates ranging from 0 to 1. Figure 10 shows a sample of the target images with the highest and lowest number of hits (our proxy for visibility).

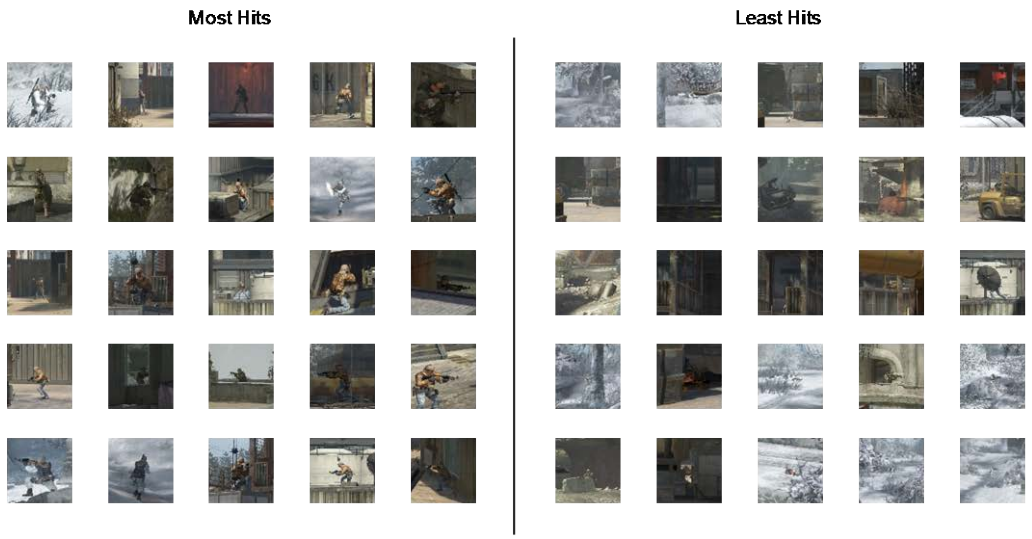

Fig. 10 Target images (ROIs) sorted by average hit rate: images with the highest number of hits (left) and images with the lowest number of hits (right)

Since target visibility affects both the hit rate and reaction time, it is reasonable to assume that there would also be an effect on classifier score. The binary classifier used in this project employs a linear discriminant function to classify the neural response elicited by each image (Touryan et al. 2010). Images with corresponding scores above zero are considered targets while images with scores below zero are considered nontargets. However, this continuous-valued score is also a measure of the strength of the object categorization response (or P300). In addition, this single score can be turned into a waveform by convolving the weight matrix, or discriminant function, with the elicited response around a finite temporal window.

$$S(\tau) = \sum_T \sum_N r(t-\tau, n) w(t, n). \quad (4)$$

Here r represents the neural response that is a function of both time (t) and EEG channel number (n). The continuous score waveform is calculated around a temporal lag (τ) by convolving the weight matrix within a lagged response window. Figure 11 shows an example of this score waveform for one subject in the 5-Hz RSVP paradigm. In this figure, the score waveform for each target is sorted by reaction time (RT) to illustrate the strong relationship between neural and behavioral response. On the left is the score waveform for each target image sorted by RT. Negative time values indicate short response latency. The plot on the right shows the relationship between lag index and RT. Lag index is the temporal lag of the peak in the score waveform. On average, the peak of the score waveform is centered at a lag of zero. However, some target images clearly result in a short latency or rapid response while others are significantly delayed in time. These large temporal dynamics can lead to a misclassification of the elicited response when only considering the score at time zero (data not shown). To quantify the amount of temporal variability, we identify the lag of the peak in the score waveform and refer to this as the "lag index".

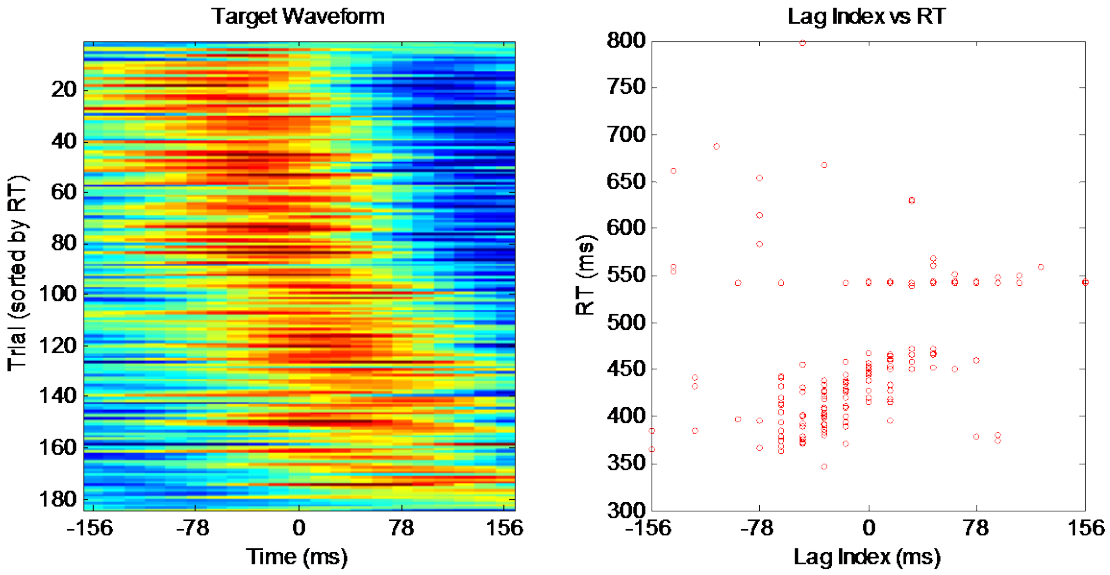

Fig. 11 Temporal dynamics of the classifier score

Over the population, we find a strong relationship between the score, lag index, and target visibility. Figure 12 illustrates this relationship by plotting the visibility of each target image as a function of the average score and the average lag index over the population (N = 13). The top plot shows visibility as a function of average classifier score. On the bottom, visibility is plotted as a function of lag index. The inset shows average score waveform sorted by average reaction time. As expected, the score monotonically increases with visibility. However, the relationship between the lag index and score is more complicated. Here, early and average latency responses have about the same visibility ratings while the longer latency responses tend to be associated with low-visibility targets. This result demonstrates the critical role of temporal dynamics in the neural response. By incorporating time in the classification of the neural response, not only is accuracy improved, but perceptual difficult can be quantified on a single-trial basis.

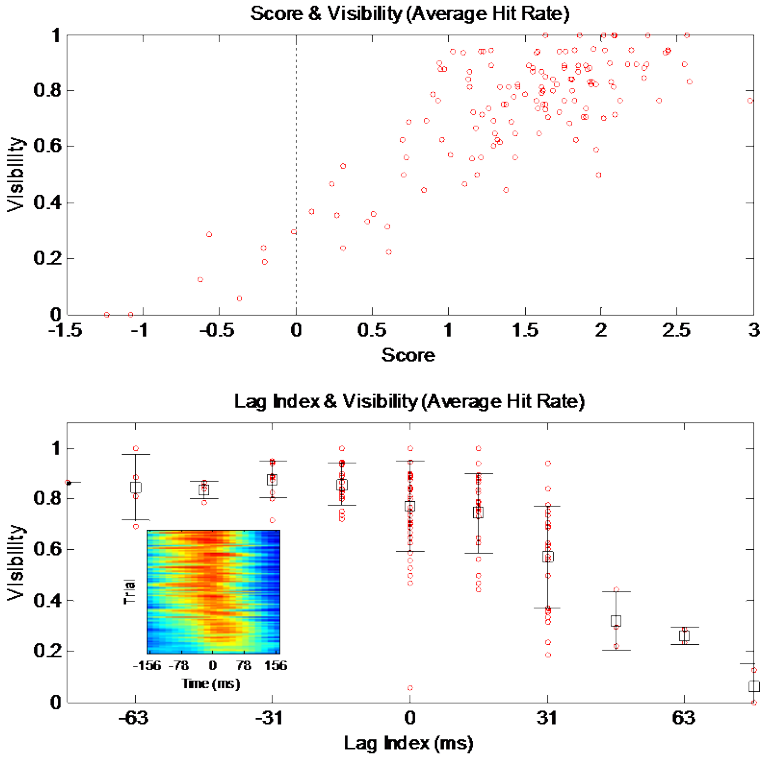

Fig. 12 Visibility and classifier score over the population

2.4 Study 4: Simulator Experiments

In addition to the RSVP experiments, we also conducted 2 studies using the Scientific Applications International Corporation (SAIC) crew station simulator (see Fig. 13). Briefly, the task is a simulated patrol of an urban environment. The MGV is driven by the computer, but the commander (experimental subject) must perform several tasks as the vehicle navigates through the urban environment. The primary task is visual target detection to identify threats. At each intersection, the vehicle stops and the subject searches for the target (in this case, a dismount carrying a gun). At half of the intersections the search is via the controllable portal while in the other half the search is via an RSVP sequence of prefiltered images. The parameters of the RSVP search component are very similar to the RSVP experiment previously described. Presentation rates of both 2 Hz and 5 Hz were used in the RSVP component. In addition to the intersection search, the subject must perform 2 other tasks: 1) identify potential improvised explosive devices near the roadside (e.g., trash bags, boxes, tires) while the vehicle is moving and 2) respond to specific radio communications.

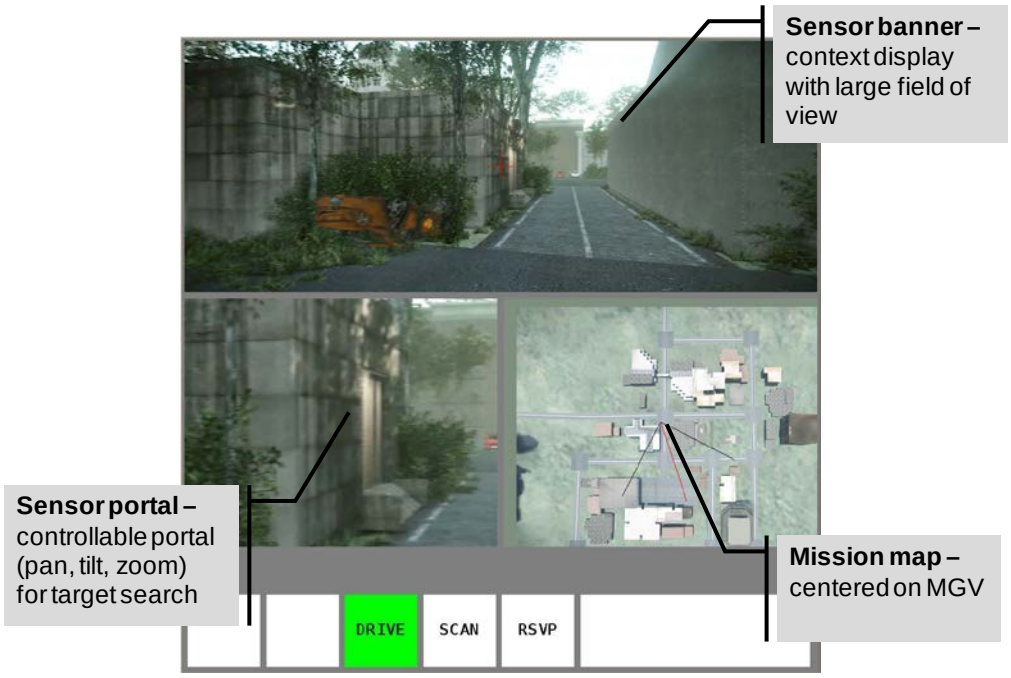

Fig. 13 Display from the SAIC simulator of a crew station commander's view. At predefined locations along the route the simulator initiates a search task either through the controllable portal or intelligent RSVP.

The overall purpose of this simulation is to integrate the real-time classification of neural signals into a complex multitasking environment and compare performance in this environment with a baseline behavioral condition. Here the neural response to each image chip in the RSVP is scored, and the top 3 chips from each intersection search are shown to the subject for confirmation. The baseline, or manual search condition, quantifies how long the subject takes to find the target with the controllable portal. In this way both the accuracy of the classifier and its impact on system performance can be quantified in a more realistic environment. A small number of participants were tested using a 2-Hz presentation rate and a larger group of 14 subjects were tested with a 5-Hz presentation rate, with some subjects run (in different sessions) at both 2 Hz and 5 Hz. For each condition, the subjects participated in a separate RSVP session prior to the simulator experiment. Individualized classification models were built from EEG data collected in these sessions and applied during the simulation runs. Results from these experiments indicate that the accuracy of the single-trial classifier is sufficient to find the target at each intersection even with the increased presentation rate.

Analysis of the 5 subjects with the 2-Hz presentation rate show that the difference in mean time to find the target (portal search: $\mu = 0.41$, $\sigma = 0.41$; RSVP: $\mu = 0.45$, $\sigma = 0.21$) is not statistically significant [$t(4) = -0.61$, $p = 0.54$]. The difference mean accuracy (portal search: $\mu = 0.80$, $\sigma = 0.4$; RSVP: $\mu = 0.92$, $\sigma = 0.27$) is also not statistically significant [$t(4) = -1.7$, $p = 0.086$]. Figure 14 shows the mean values by subject.

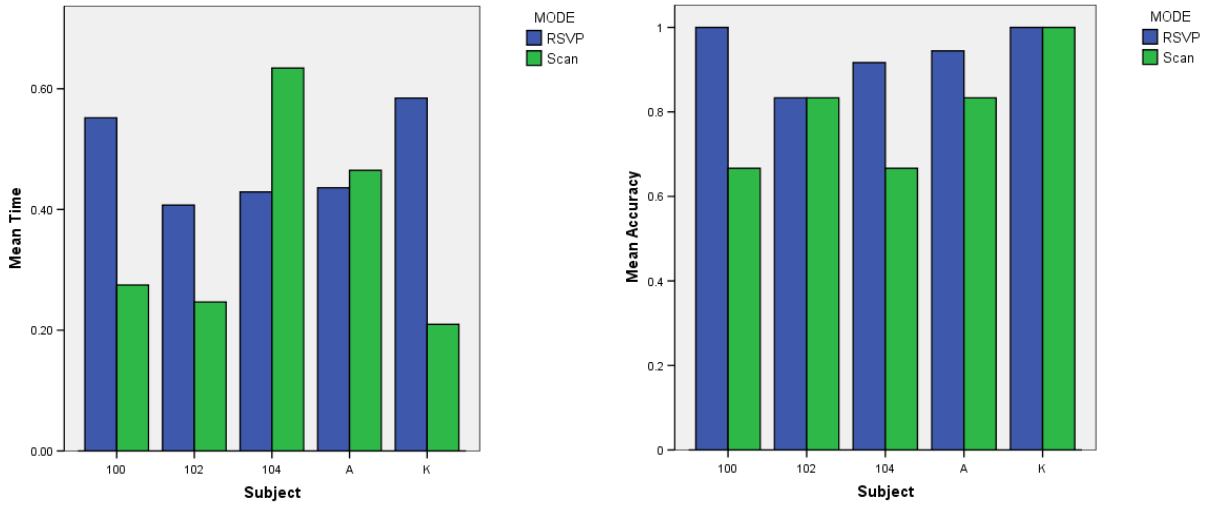

Fig. 14 Simulator results for 5 subjects with 2-Hz RSVP presentation rate. Graphs compare time (in minutes) to find target and accuracy for RSVP vs. portal search.

Results for the 14 subjects with the 5-Hz presentation rate are given in Fig. 15. At the faster presentation rate the difference in mean time to find the target (portal search: $\mu = 0.64$, $\sigma = 0.49$; RSVP: $\mu = 0.23$, $\sigma = 0.11$) is significant [$t(13) = 7.8$, $p < 0.001$] while the difference in mean accuracy (portal search: $\mu = 0.80$, $\sigma = 0.4$; RSVP: $\mu = 0.85$, $\sigma = 0.35$) is not statistically significant [$(t(13) = -1.13$, $p = 0.259$].

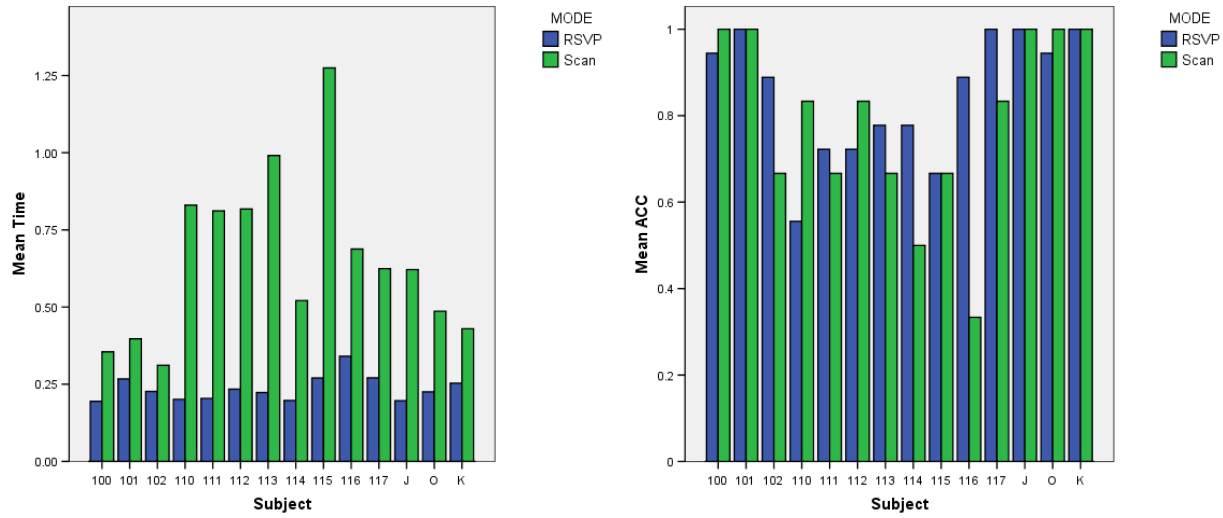

Fig. 15 Simulator results for 14 subjects with 5-Hz RSVP presentation rate. Graphs compare time (in minutes) to find target and accuracy for RSVP vs. portal search.

3. Alternate RSVP Task Experiments

In parallel with the simulator studies, we conducted 3 experiments that focused on novel paradigms for using RSVP in the MGV crew station. These studies attempted to qualify and understand the type of performance that could be achieved by using a target detection system based on the analysis of brain responses. The goal of the first experiment was to evaluate the performance of single-trial detection during a dual-task paradigm. The second experiment was aimed at a novel method for improving the accuracy of target detection by measuring 2 responses to the same target within the constraints of the real-time neural classification. The third experiment examined the performance of a collaborative target detection system in which several observers are involved in the same target detection task, but each person has a different angle of observation of the potential target. Ten healthy subjects have participated in the first 2 main experiments (M = 20 years old, standard deviation [SD] = 1). The last experiment was carried out with a team of 5 healthy subjects (M = 20 years old, SD = 1.4).

In each experiment, realistic images were presented to the subjects. The visual stimuli set consisted of 683- × 384-pixel color images. These images were taken from “Insurgency: Modern Infantry Combat” (New World Interactive, Denver, CO, 2010–2014), a total conversion modification of the video game “Half-Life 2” (Valve Corporation, Bellvue, WA, 2004). These images were selected because of the high degree of similarity with the environments used in the TARDEC and US Army Research Laboratory crew station simulators. Images of the visual stimuli are presented in Fig. 16.

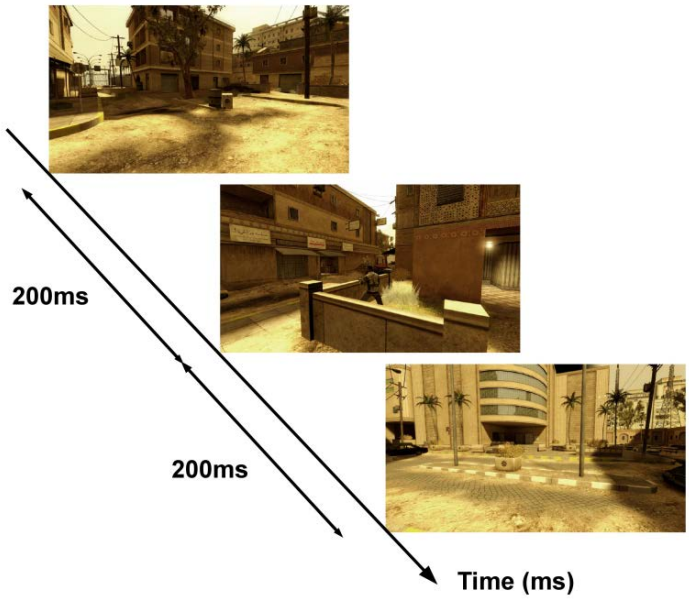

Fig. 16 Example of images presented to the user during the experiments

3.1 Effect of a Dual-Task Condition with Visual Tasks

The goal of the first experiment was to evaluate the impact of more realistic dual-task conditions on neural classification. Previously (see year 1 results) we have shown that there is a decrease in both behavioral and classification performance with the dual-task condition. In those experiments, the 2 tasks were in different sensory modalities (visual and auditory). In the simulator environment, operators have to deal with multiple visual tasks, so here we focused on using 2 visual tasks for our study.

Study 5: Two RSVP Tasks

In the first experiment we used 2 RSVP tasks each with an image presentation frequency of 5 Hz and a target probability of 0.1. For each task the goal is to detect the presence of a person in a scene. For both tasks we only consider the neural response for the detection of targets. To be able to detect targets in both tasks at the same time, the images in both RSVP tasks are not presented in phase. This means that the 2 images are never presented at the same time. Therefore, it is possible to detect the targets in both tasks independently.

Two conditions were tested: single task and dual-task. The mean area under the ROC curve across subjects for the single task condition is 0.796 ± 0.025 while the mean is only 0.689 ± 0.020 for the dual-task condition. This decrease in performance is statistically significant [$t(9) = 6.315$, $p < 0.001$] and highlights the difference between single and dual-task. Despite this decrease, the user views twice as many images.

3.2 An RSVP Task and a Behavioral Task

In the second experiment, observers were presented with an RSVP task on the left of the screen and a map task on the right of the screen. The map task consisted of pressing a key on the keyboard when a green dot was presented on the map. A display of the visual stimuli is depicted in Fig. 17. Three conditions were tested to evaluate the impact of the dual-task condition: the RSVP task only, the behavioral task only, and both tasks simultaneously. For the behavioral task only, the hit-rate and the precision were 91.3% and 97.2%, respectively. This level of performance is not surprising, as the task was easy. For the RSVP task only, the mean area under the ROC curve across subjects was 0.837. When both tasks are performed simultaneously, the performance of the behavioral task drops. The hit rate and precision were 86.6% and 92.3%, respectively. There was a significant difference between the single and dual-task condition for the behavioral task ($p < 0.05$). However, the mean area under the ROC curve of the RSVP task was 0.838. In this case, there was no statistical difference between the single and the dual-task condition.

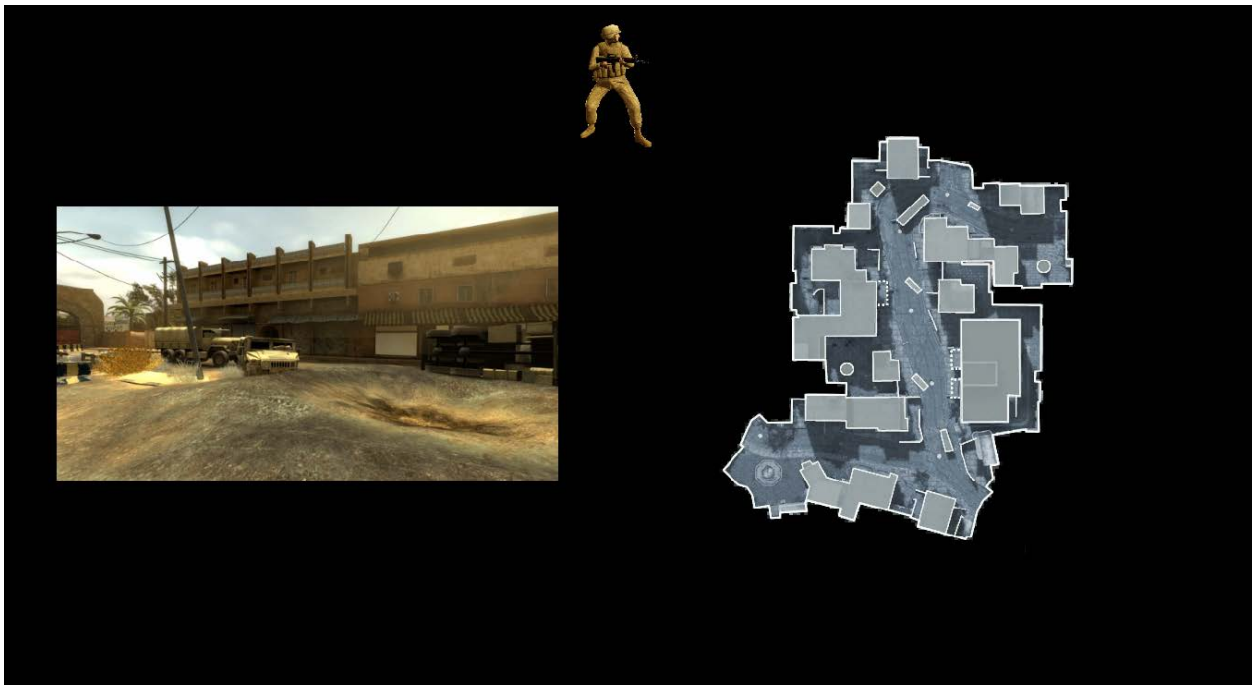

Fig. 17 RSVP task (neural detection) (left) and behavioral task (right)

These results suggest that a decrease in performance can be expected when the user is engaged in several tasks. In the first experiment both tasks were identical and posed the same difficulty. In the second experiment, tasks were different: The behavioral task was easier.

3.3 Improvement of the RSVP Paradigm

For target detection to occur in real time it is impossible to repeat the presentation of images on the screen. Therefore single-trial detection should be used for target detection. With only one trial it is often difficult to obtain reliable results due to the poor signal-to-noise ratio (SNR) of the signal. We propose a new paradigm where the constraint of the images occurring in real time is preserved. This paradigm is composed of 2 RSVP streams. These 2 streams of images are identical, the only difference being that the second one is delayed in time. If a target appears in the first (primary) RSVP stream, this target will be presented later in the second stream. The subject pays attention to the primary, real-time RSVP stream and then, if they detected a target, switch their attention to the second stream to confirm the presence of a target previously seen in the primary stream. After switching for the confirmatory presentation, the subject then switches back to the primary RSVP stream. With this strategy 2 ERPs in the EEG signal are produced for the same visual stimulus.

For single-trial detection of both RSVP streams, the mean area under the ROC curve across subjects was 0.805. As there may be a difference between the ERP evoked by the 2 RSVP streams, the detection was analyzed separately for each RSVP stream. For the primary RSVP stream, where the subject sees the target for the first time, the mean area under the ROC curve

is 0.818. The mean area under the ROC curve for the second RSVP stream, where the subject confirms the presence of a target, is 0.795. This difference in area under the ROC curve between the 2 streams is not statistically significant ($p = 0.057$). However, the combination of these 2 trials improved the accuracy of the target detection, increasing the mean area under the ROC curve to 0.873. The ROC curve of each subject after the combination of 2 trials is presented in Fig. 18. With this paradigm, the mean area under the ROC curve is increased by combining 2 trials while keeping the presentation of the visual stimuli in real time.

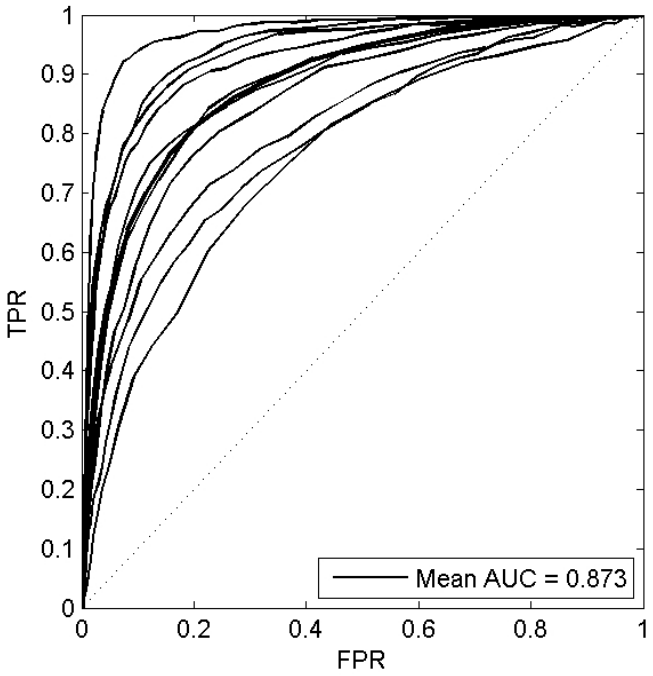

Fig. 18 ROC curves for each subject after the combination of 2 trials

3.4 Collaborative BCI for Improving Overall Performance

Whereas classical neural detection systems are based only on the response of a single individual, the combination of the EEG signals from several individuals can improve the overall accuracy (Eckstein et al. 2012). Indeed, combining trials improves the SNR of the EEG signal. Averaging several trials over time has been done since the early days of BCI with the P300 speller. The main challenge is to find applications where multiple trials are natural or inherent in the task. Since the combination of trials from several subjects is known to increase the SNR, we would like to consider BCI paradigms that require several subjects and where the underlying task is identical across subjects.

We examine a collaborative BCI where different subjects are involved for the detection of the same targets at the same time. These subjects observe the same sequence of target and nontarget objects and scenes but from different viewpoints. Each subject has a different physical position

in the environment so each subject has a different view of the target. Figure 19 shows an example of the same target viewed from 5 different positions. The goal of this paradigm is to enhance the target detection accuracy by combining the neural responses of subjects who are doing the same task. This paradigm is composed of an RSVP task that contains realistic images. The images of targets correspond to a view from each of 5 angles around the target, as if 5 different observers encircle the target. For single-trial detection the average area under the ROC curve across subjects is 0.887. The ROC curve of each subject is presented in Fig. 20. With a weighted average combination of the different outputs from each subject, the area under the ROC curve is 0.991. These promising results show the possibility of reliably detecting targets in real time by combining the results for several subjects doing the same global task. The performance based on the area under the ROC curve is presented in Fig. 21 as a function of the number of subjects involved in the decision.

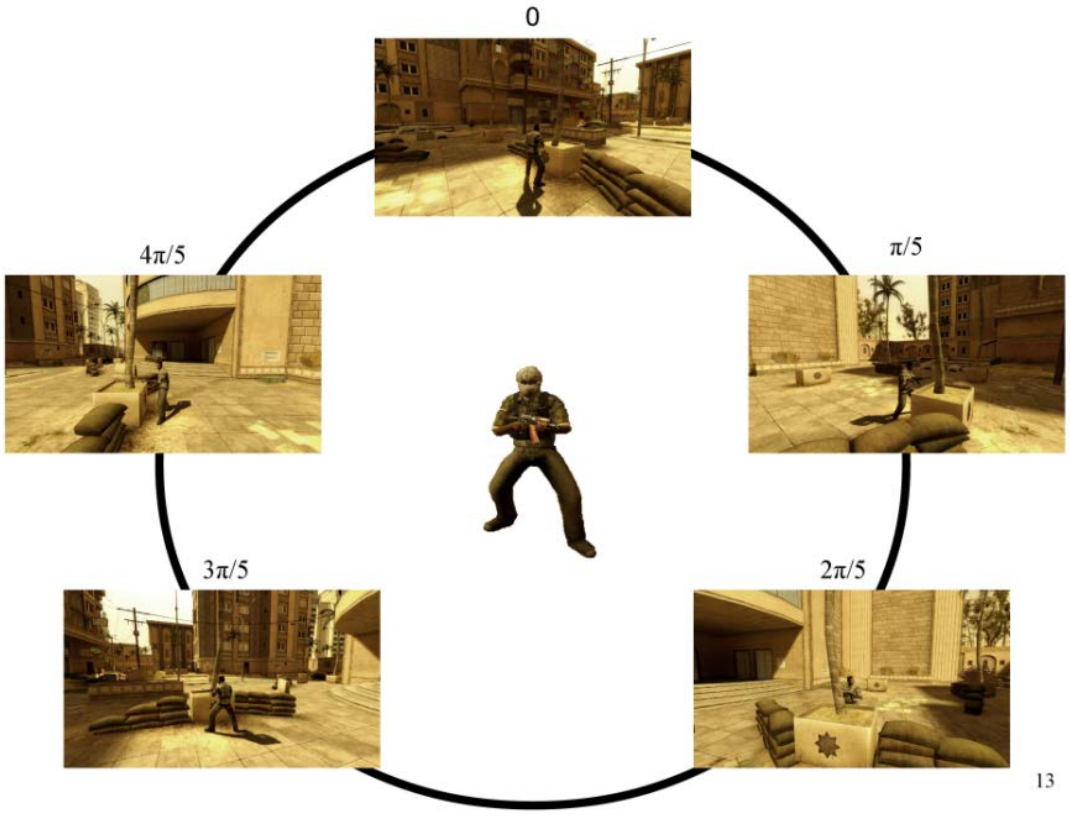

Fig. 19 Visual stimuli from 5 angles that are observed by 5 different subjects

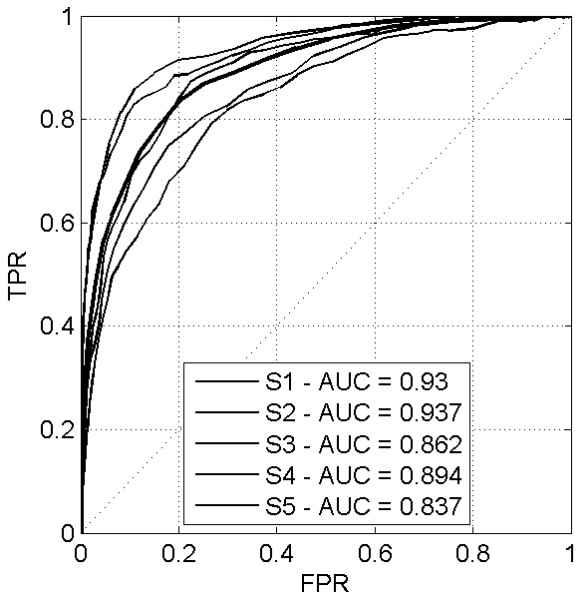

Fig. 20 ROC curves for each subject

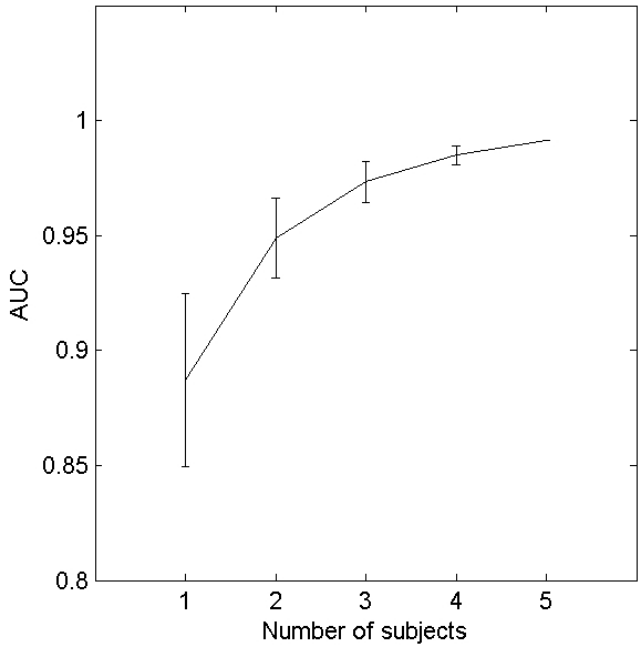

Fig. 21 Area under the ROC curve as a function of the number of observers

This set of experiments shows the effect of a dual-task condition in a realistic setting. It demonstrates 2 paradigms for improving target detection based on the detection of neural signals. With the combination of 2 trials, an improvement in the area under the ROC curve from 0.805 to 0.873 was achieved. With the combination of trials across 5 subjects we were able to achieve an area under the ROC curve of 0.991, i.e., an almost perfect performance.

4. Multiclass Classification of Neural Signals

In the studies with RSVP and neural classification described previously and in year 1, each image (or video clip) presented either contains a target or it does not. Current systems based on the detection of neural signatures look for a single type of response to detect. Hence the classification methods that are considered in these systems are binary classifiers (target versus nontarget). In operational settings, there can be several classes of images to which an operator may respond in different ways. For example, there may be images that contain only the background environment. Others may contain noncombatant civilians. Images that contain insurgents or threats constitute a third type.

Study 6: Multiclass Discrimination

During this year's work, a study was carried out to look at methods for classifying an operator's neural response to an image or video clip into more than the 2 classic target and nontarget categories. This investigation of multiclass classification of single-trial ERPs during a rapid serial visual presentation task used short video clips (see Fig. 22). Each trial contained potential targets that were human or nonhuman, stationary or moving. The goal of the classification analysis was to discriminate between 3 classes: a moving target human (MTH), a moving target nonhuman (MTNH), and a nonmoving target human (NMTH).

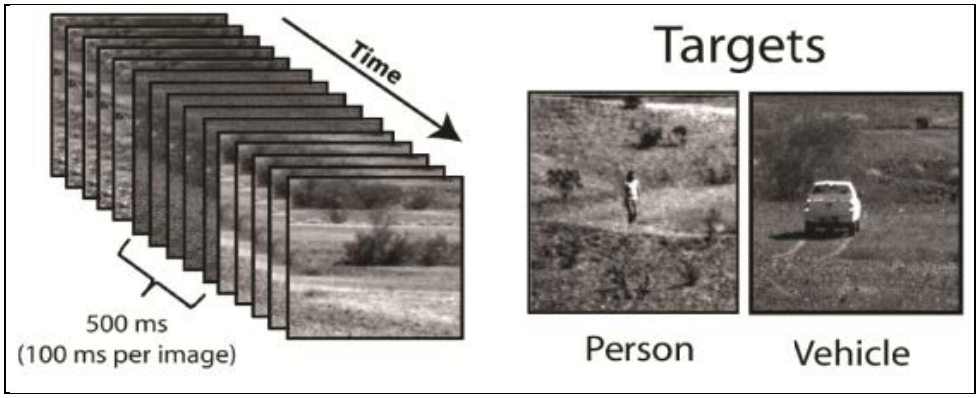

Fig. 22 RSVP task and examples of targets

The binary classification of each class with a one-versus-all approach was first evaluated. The area under the ROC curve for these binary classifications is presented in Fig. 23. The mean area under the ROC curves for the detection of an MTH, MTNH, and NMTH was 0.907, 0.855, and 0.914, respectively. The detection of an MTH is easier than an MTNH ($p < 0.05$, $t = 2.404$). Detection of an MTH was better than both an MTNH ($p < 0.05$, $t = 2.589$) and an MTH ($p < 0.05$, $t14 = 2.589$). These results suggest that it is easier to detect stationary human targets.

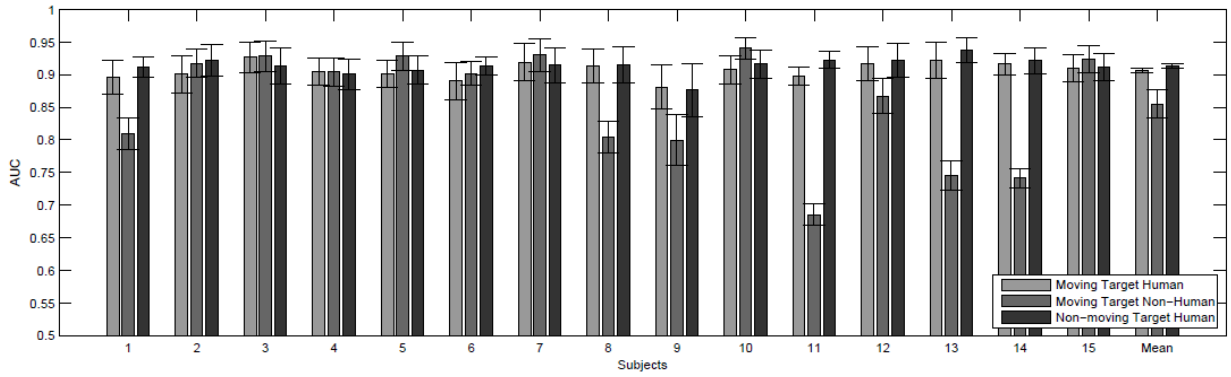

Fig. 23 Area under the ROC curve for the binary classification. The error bars correspond to the standard error across sessions for each subject and across subjects for the mean.

For the multiclass classification, we consider the argmaximum of the outputs from the different binary classifiers. The performance of the multiclass classification can be represented as an ROC surface by weighting the decision of each binary classifier (equivalent to a threshold for a binary classifier). The resulting ROC surface represents the performance for all the classes for different sets of weights. (Ferri et al. 2003, Landgrebe and Duin 2007)

The analysis revealed that a mean volume under the ROC surface of 0.878 (see Figs. 24 and 25). These results suggest that it is possible to efficiently discriminate between more than 2 types of evoked responses using single-trial detection.

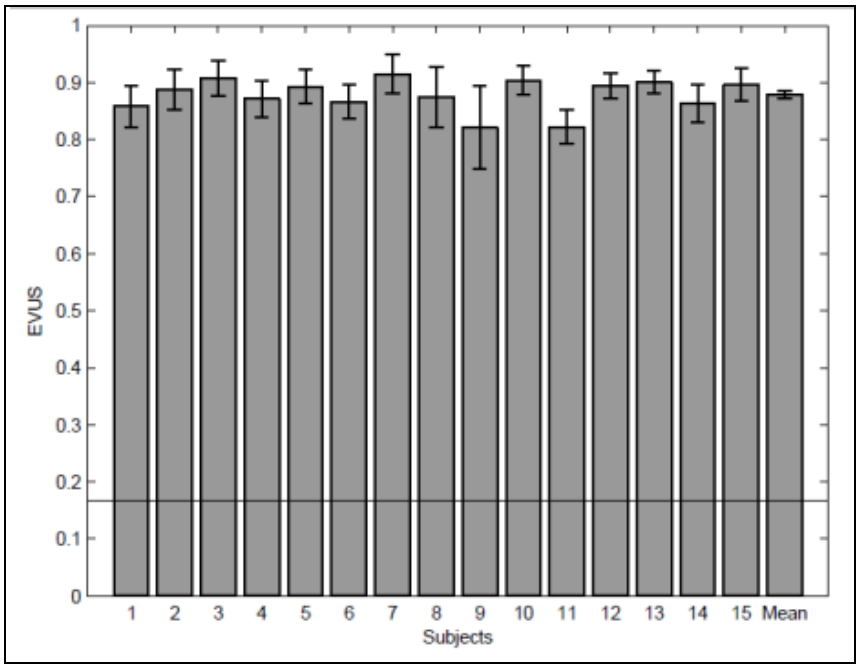

Fig. 24 Estimated volume under the surface (EVUS) for each subject. The error bars correspond to the standard error across sessions for each subject and across subjects for the mean.

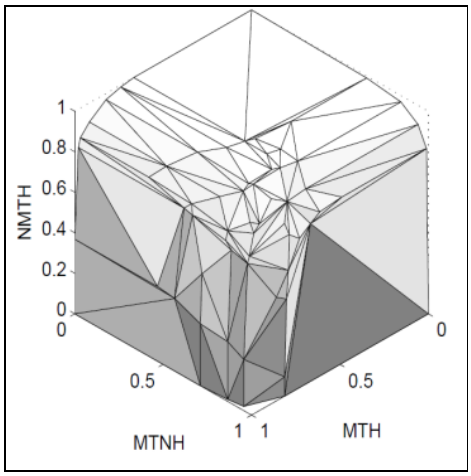

Fig. 25 Example of an ROC surface representing the performance of subject 1 (EVUS = 0.9507)

In Fig. 26 the grand-averaged ERP waveforms for each stimulus class are plotted with a baseline correction of –200 to 0.0 ms on the electrodes Fz, Cz, Pz, Oz, P7, and P8. These plots were created for each stimulus class and low pass filtered at 30 Hz. Continuous artifact-free data were time-locked to stimulus onset and epoched from –200 to 1,000 ms. Only targets followed by a response within 200–1,000 ms or nontargets followed by no response were included in the analysis.

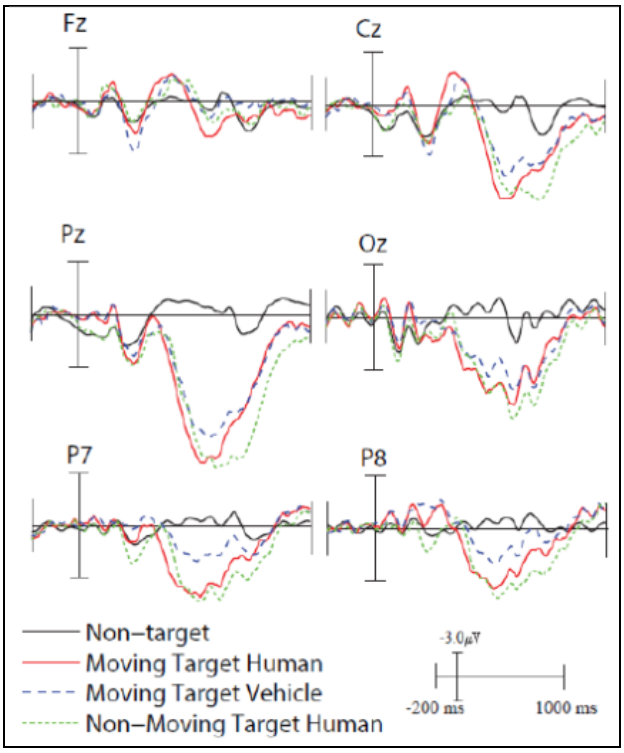

Fig. 26 Grand-averaged ERP waveforms for each stimulus

5. Predicting Performance

A small study was conducted using data from year 1 to look at techniques for predicting performance from EEG with RSVP tasks. A number of recently published studies have demonstrated that perceptual and attentional performance can be predicted by the amplitude and phase of oscillatory substrates in the brain. In particular, prestimulus alpha power and phase have been repeatedly shown to be predictive of whether subjects will detect or miss otherwise perceptually identical stimuli. Based on this work, we explored the spatial and temporal characteristics of this oscillatory activity within our RSVP tasks. Our tasks are particularly well-suited to investigating this issue because we are able to investigate the relationship between brain activity and performance on the single trial level as well as for longer periods of time. Specifically the task is divided into 50 blocks of 240 images, and each block of 240 images (each 2 min long) was divided into miniblocks of 10 images (each 5 s long). In the year 1 studies, we used images of faces and cars as stimuli. We focused on trial averaging over both the blocks and miniblocks within the RSVP task in which the probability of a face target was 0.5. The mean behavioral performance (percent correct, n = 8) across the blocks is shown in Fig. 27. It is clear from this figure that there are systematic fluctuations in performance across the blocks and that performance is highly variable.

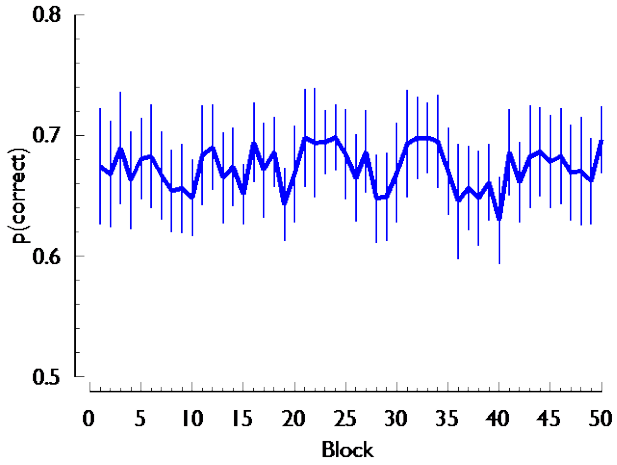

Fig. 27 Behavioral performance

Study 7: Predicting RSVP Performance

To investigate the characteristics of the EEG signal that can discriminate between these periods of good and bad performance, we performed 2 separate analyses. Based on previous work showing that fluctuations in occipital alpha can discriminate between hit and miss trials, we divided the miniblocks into those in which all the targets were correctly detected (hit blocks) and those in which all the targets were missed (miss blocks). Then we coupled the power spectral

density in the alpha frequency band at occipital electrodes (PO3/4, O1/2, Oz) for the 40 s prior to those hit and miss blocks. The results of this analysis are shown in Fig. 28. The key finding from this analysis was that there was significantly more alpha at occipital electrodes 10 s prior to a miss block than a hit block [$t(7) = 2.43$, $p < 0.05$].

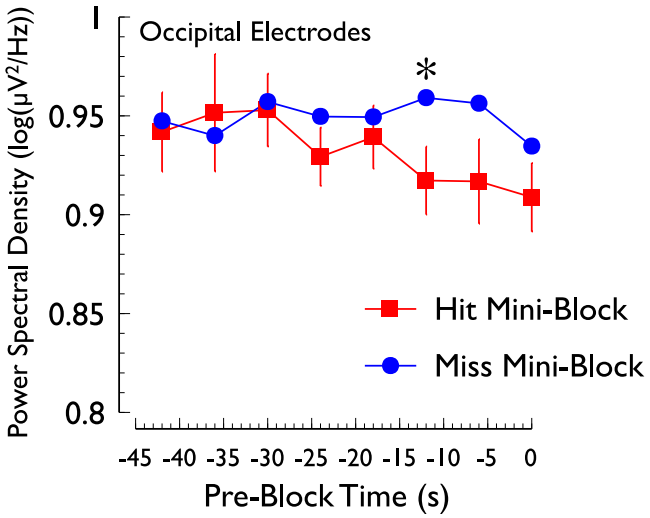

Fig. 28 Mean power spectral density in the alpha frequency band

The second analysis we performed divided all the blocks into those in which performance was the best and those in which performance was the worst (relative to the median across blocks). The best and worst blocks were then compared in terms of the oscillatory activity induced by the RSVP stream itself, otherwise known as the steady state visually evoked potential (SSVEP). To compute the SSVEP we band pass filtered the data for each miniblock (best or worst) centered on the stimulation frequency of 2 Hz and averaged the resulting wave forms at the same occipital electrodes used in the first analysis (see Fig. 29A). Visual inspection of these waveforms clearly indicates that the amplitude of the SSVEP was higher during blocks in which performance was low. These amplitude differences were quantified by computing the mean peak-to-peak amplitude for each type of block. The results of this analysis, shown in Fig. 29B, revealed that SSVEP amplitude was significantly higher in best-performing blocks ($p < 0.007$). Together these findings are consistent with recent studies linking increases in occipital alpha to reductions in behavioral performance and suggest that these fluctuations play a key role in the spatio-temporal dynamics of attention.

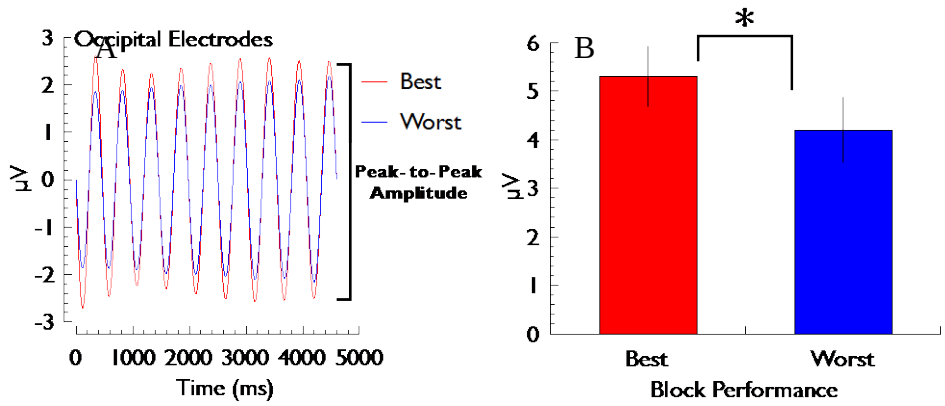

Fig. 29 Mean SSVEP amplitude at 2 Hz measured at electrodes PO3/4 and O1/2/z: A) mean across best and worst performing mini blocks and B) mean peak-to-peak SSVEP amplitude

6. Conclusions

This year the team made significant progress in developing a simulation environment to test the performance of state-of-the-art neural classification techniques in an operational context. In the previous year we focused on determining the optimal parameters for classification of the neural response in an RSVP paradigm. The parameters investigated included target presentation properties (e.g., size, eccentricity, and rate), the effects of changes in attentional state on classification accuracy, and the effect of operator multitasking on system performance. This year, we focused on the specific application of the automated neural processing to a US Army-relevant system. Our intent was to replace the manual visual search task currently used to both identify targets and maintain situational awareness in MGVs. Specifically, the RSVP paradigm in combination with automated classification of the neural response would replace the manual control of an imaging sensor on the MGV. Therefore, instead of an operator manipulating the PTZ camera to scan the environment, images of the vehicle's surroundings containing potential targets would be rapidly presented and subsequently sorted based on the operator's neural response. The operator could then review the most relevant images for target confirmation.

This second stage of development consisted of 2 elements. First we sought to quantify the potential tradeoff of replacing a manual search with RSVP. To accomplish this we conducted an experiment to compare the time-to-target and accuracy of these 2 paradigms. Secondly we developed a simulation environment based on the MGV crew station. This simulator was designed to switch between the 2 search paradigms and was fully integrated with a real-time EEG processing system. In addition, the simulator incorporated multitasking aspects of the crew

station including auditory and text communications. Together these results demonstrate the feasibility and potential benefits of integrating automated neural processing technology into Army systems.

In the final year of this project we will build on the study results and software that has been developed to create a state-of-the-art, standalone, real-time, RSVP-based system for target detection. In addition, the simulation environment that was built this year will be further developed into a flexible multitasking system called the RSVP-based Adaptive Virtual Environment with Neural-processing (RAVEN). Both RAVEN and the stand-alone system will support prototyping and evaluation of neural processing in operational Army applications. The key results of our studies are summarized in Table 3.

Table 3 Summary of key year 2 results

Result	Study
Searching the environment for threats using RSVP gave significantly higher accuracy than a manually controlled scan. The mean accuracy for manual portal search was 0.85 while for RSVP it was 0.99.	1
With a slew-to-cue function the initial accuracy of the portal position does not strongly correlate with search time. Untrained operators tend to follow a contextual search.	1
Portal speed (PTZ speed) did not significantly affect search time.	1
Participants significantly reduced their search time from the first to the second session.	1
There was no significant correlation between target salience and search time.	1
A measure of temporal displacement of the neural response to a target called the lag index was defined. Indices that represent longer, delayed responses are associated with low-visibility targets. This measure could be useful to improve classification accuracy and to quantify perceptual difficulty.	3
Using a 2-Hz presentation rate for RSVP, there was no significant difference in either accuracy or speed in finding targets. With a 5-Hz presentation rate, the accuracy was not significantly different but operators found the target more quickly with RSVP than with portal search.	4
Comparing neural classification when a subject simultaneously views 2 RSVP streams at 5 Hz with target probability of 0.1 to a single stream, we found that classification accuracy decreased from 0.86 to 0.75 AUC.	5
When simultaneously performing RSVP and a behavioral task, accuracy on the behavioral task degraded significantly. The classification of the neural response to RSVP under this dual task paradigm did not differ significantly from RSVP alone.	5
Presenting images side-by-side in 2 identical RSVP streams with one delayed in time increases the overall accuracy of the neural classification when the classification of the responses to the separate streams are combined.	5
Combining the neural responses of 5 subjects viewing the same target from different viewpoints results in improved classification accuracy.	5
By combining the output of binary classifiers, it is possible to discriminate between the neural responses to more than 2 types of targets and nontargets.	6
Examining the power spectra of the EEG signal prior to RSVP single trials, we can predict whether or not a subject will detect or miss a target with some degree of accuracy.	7
Comparing the blocks of RSVP in which the subjects performed best with those in which they performed worst, we found that SSVEP amplitude was significantly higher with good performance.	7

7. References

Eckstein MP, Das K, Pham BT, Peterson MF, Abbey CK, Sy JL, Giesbrecht B. Neural decoding of collective wisdom with multi-brain computing. Neuroimage. 2012;59(1):94–108.

Fawcett, T. An introduction to ROC analysis. Pattern Recognition Letters. 2006;27:861–874.

Ferri C, Hernández-Orallo J, Salido MA. Volume under the ROC surface for multiclass problems. In: Lavrač N, Gamberger D, Blockeel H, Todorovski L, editors. ECML 2003. Proceedings of the 14th European Conference on Machine Learning; 2003 Sep 23–27; Cavtat-Dubrovnik, Croatia. Berlin (Germany): Springer-Verlag; c2003. p. 108–120.

Gibson L, Touryan J, Ries AJ, Cecotti H, Giesbrecht B. Adaptive integration and optimization of automated and neural processing systems: establishing neural and behavioral benchmarks of optimized performance. Aberdeen Proving Ground (MD): Army Research Laboratory (US); 2012 Jul. Report No.: ARL-TR-6055. Also available at: http://www.arl.army.mil/arlreports/2012 /ARL-TR-6055.pdf.

Itti L, Koch C. A saliency-based search mechanism for overt and cover shifts of visual attention. Vision Res. 2000;40:1489–1506.

Landgrebe TC, Duin RP. Approximating the multiclass ROC by pairwise analysis. Pattern Recognition Letters. 2007;28:1747–1758.

Luo A, Sajda P. Comparing neural correlates of visual target detection in serial visual presentations having different temporal correlations. Front Hum Neurosci. 2009;3:5.

Sajda P, Gerson A, Parra L. High-throughput image search via single-trial event detection in a rapid serial visual presentation task. Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering; 2003 Mar 20–22; Capri Island, Italy. New York (NY): Institute of Electrical and Electronics Engineers; c2003. p. 7–10.

Sajda P, Pohlmeyer E, Wang J, Parra L, Christoforou C, Dmochowski J, Hanna B, Bahlmann C, Singh MK, Chang SF. In a blink of an eye and a switch of a transistor: cortically coupled computer vision. Proceedings of the IEEE. 2010;98:462–478.

Torralba A, Murphy K, Freeman W, Rubin, M. Context-based vision system for place and object recognition. Proceedings of the IEEE. 2003;271:273–280.

Torralba A, Oliva A, Castelhano M, Henderson J. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. Psychological Review. 2006;113:766.

Touryan J, Gibson L, Horne J, Weber P. Real-time classification of neural signals corresponding to the detection of targets in video imagery. Proceedings of International Conference on Applied Human Factors and Ergonomics 2010; 2010 17–20 July; Miami, FL. Boca Raton (FL): CRC Press; c2010. p. 60.

List of Symbols, Abbreviations, and Acronyms

ABM	Advanced Brain Monitoring
AUC	area under the ROC curve
BCI	brain-computer interface
EEG	electroencephalogram
ERP	event-related potential
EVUS	estimated volume under the surface
FOV	field of view
Hz	hertz
MGV	manned ground vehicle
MTH	moving target human
MTNH	moving target non human
NMTH	nonmoving target human
PPV	positive predicted value
PTZ	pan tilt zoom
RAVEN	RSVP-based Adaptive Virtual Environment with Neural-processing
ROC	receiver operating characteristic
ROI	region of interest
RSVP	rapid serial visual presentation
RT	reaction time
SAIC	Science Applications International Corporation
SNR	signal-to-noise ratio
SSVEP	steady state visual evoked potential
TARDEC	Tank and Automotive Research, Development and Engineering Center
TPR	true positive rate

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
2 (PDF)	DIRECTOR US ARMY RESEARCH LAB RDRL CIO LL IMAL HRA MAIL & RECORDS MGMT
1 (PDF)	GOVT PRINTG OFC A MALHOTRA
1 (PDF)	DIR USARL RDRL HRS C A RIES

INTENTIONALLY LEFT BLANK.